

# Karakterizacija likova u dječjim pričama

## Diplomski rad

**Gorana Levačić**

**Voditelj rada: izv. prof. dr. sc. Saša Singer**

Prirodoslovno matematički fakultet - Matematički odsjek

28. studenog 2016.

- 1 Obrada prirodnog jezika
- 2 Prepoznavanje imenovanih entiteta
- 3 Razrješavanje koreferencije
- 4 Analiza sentimenta
- 5 Razvoj aplikacije za karakterizaciju likova
- 6 Rezultati
- 7 Zaključak

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

# Obrada prirodnog jezika

# Obrada prirodnog jezika

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

Cilj diplomskog rada – razviti sustav koji će:

- 1 prepoznati likove u dječjim pričama
- 2 za svakog lika zaključiti je li dobar ili loš

⇒ problem obrade prirodnog jezika

# Obrada prirodnog jezika

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

## Objedinjuje:

- računarstvo
- umjetnu inteligenciju
- računalnu lingvistiku

## Problemi:

- razumijevanje prirodnog jezika
- strojno prevođenje
- prepoznavanje govora
- generiranje teksta
- ⋮

# Obrada prirodnog jezika

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

Problemi prepoznati u problemu karakterizacije likova:

- prepoznavanje imenovanih entiteta
- razrješavanje koreferencije
- analiza sentimenta

# Obrada prirodnog jezika

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

Pristup rješavanju problema obrade prirodnog jezika:

- 1 formalno modeliranje
- 2 statističke metode
- 3 strojno učenje

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

# Prepoznavanje imenovanih entiteta



# Prepoznavanje imenovanih entiteta

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

- prepoznavanje i klasificiranje izraza u tekstu koji se odnose na imenovane entitete

Imenovani entiteti:

- osobe
- lokacije
- organizacije
- 

*Red Riding Hood* walked through *Grünewald* forest.

*Person* *Location*

# Prepoznavanje imenovanih entiteta

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

U problemu karakterizacije likova jedna kategorija – likovi:

- imenovani – *Little Red Riding Hood*
- neimenovani – *evil witch*

Postojeći NER sustavi prepoznaju samo imenovane likove  
⇒ treniranje vlastitog modela

# Prepoznavanje imenovanih entiteta

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

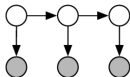
Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

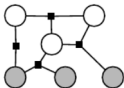
Zaključak

## Metode:

- **skriveni Markovljevi modeli** – pretpostavljaju da oznaka ovisi samo o toj riječi i o oznaci prethodne riječi u nizu



- **uvjetna slučajna polja** – za ulaz uzimaju više podataka o samoj riječi, ne promatraju nužno isključivo prethodnu oznaku



Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

**Razrješavanje  
koreferencije**

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

# Razrješavanje koreferencije

# Razrješavanje koreferencije

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

- prepoznavanje izraza u tekstu koji se odnose na isti izvanjezični entitet

*She always wore red cloak with a hood, so people called her Little Red Riding Hood.*

# Razrješavanje koreferencije

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

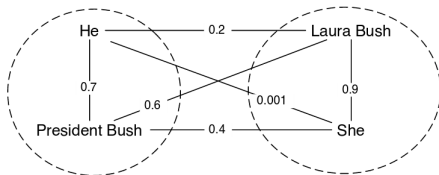
Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

## Metode:

- **model parova** – promatra vjerojatnost međusobne koreferencije pojedinih parova izraza



- **model logike prvog reda** – poopćenje prethodnog modela, prepoznaje grupe koreferentnih izraza

# Razrješavanje koreferencije

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

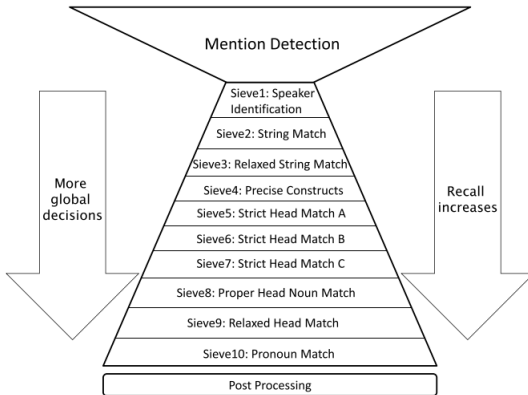
Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

Metode (nastavak):

- **metoda višeprolaznog sita** – Stanfordov *state-of-the-art* sustav



Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

**Analiza  
sentimenta**

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

# Analiza sentimenta



# Analiza sentimenta

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

Određivanje stava zadanog teksta:

- pozitivan ili negativan
- subjektivan ili objektivan
- više različitih karakteristika nekog pojma

*This book has boring plot, but great characters.*

# Analiza sentimenta

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

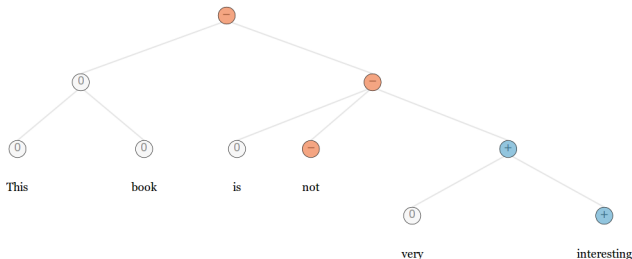
Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

## Metode:

- **"vreća riječi"** – prikaz teksta isključivo kao skupa riječi koje ga čine
- **"vreća stavova"** – promatra stavove koji se sastoje od korijena (*interesting*), modifikatora (*very*) i negatora (*not*)
- **rekurzivni duboki modeli** – temelje se na neuronskim mrežama, Stanfordov *state-of-the-art* model



Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

# Razvoj aplikacije za karakterizaciju likova

# Razvoj aplikacije za karakterizaciju likova

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

- programski jezik Java
- upotreba *Stanford CoreNLP* skupa alata za obradu prirodnog jezika
- treniranje vlastitih CRF NER modela za prepoznavanje likova
- korištenje već istreniranih modela za razrješavanje koreferencije i analizu sentimenta

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

**Rezultati**

Zaključak

# Rezultati

# Rezultati

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

## Testiranje:

- Snjeguljica
- Pepeljuga
- Vuk i sedam kozlića
- Pljačkaš zaručnik

## Rezultati:

- većinom ispravno prepoznaje likove
- dobre likove koji su sretni karakterizira kao dobre
- dobre likove koji su nesretni karakterizira kao loše
- loše likove najčešće karakterizira kao loše

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

# Zaključak

# Zaključak

Karakterizacija  
likova

Gorana  
Levačić

Obrada  
prirodnog  
jezika

Prepoznavanje  
imenovanih  
entiteta

Razrješavanje  
koreferencije

Analiza  
sentimenta

Razvoj  
aplikacije za  
karakterizaciju  
likova

Rezultati

Zaključak

- prihvatljiva uspješnost s obzirom na složenost problema

Moguća poboljšanja:

- prepoznavanjem surečenica u rečenici
- promatranjem samo onih rečenica u kojima je subjekt lik
- treniranjem zasebnog modela za analizu sentimenta

Drugi pristup rješavanju problema:

- treniranjem jedinstvenog modela nad označenim podacima
- oznake *lik*, *karakter* za pojedine dijelove teksta
- neuronske mreže